

Data Architecture concepts (Olap vs OLTP, data lakes, data warehouses, data mart)

OLAP

tujuan : OLAP membantu Anda menganalisis data dalam volume besar untuk mendukung pengambilan keputusan.

sumber data : OLAP menggunakan data historis dan agregat dari beberapa sumber.

Struktur data : OLAP menggunakan basis data multidimensional (kubus) atau relasional.

Model data : OLAP menggunakan skema bintang, skema kepingan salju, atau model analitik lainnya.

Volume data : OLAP memiliki persyaratan penyimpanan yang besar. Misalnya, terabita (TB) dan petabita (PB).

Waktu respons : OLAP memiliki waktu respons yang lebih lama, biasanya dalam detik atau menit.

Contoh penerapan : OLAP baik digunakan untuk menganalisis tren, memprediksi perilaku pelanggan, dan mengidentifikasi profitabilitas.

OLTP

tujuan : OLTP membantu Anda mengelola dan memproses transaksi waktu nyata.

sumber data : OLTP menggunakan data waktu nyata dan transaksional dari satu sumber.

Struktur data : OLTP menggunakan basis data relasional.

Model data : OLTP menggunakan model yang dinormalisasi atau didenormalisasi.

Volume data : OLTP memiliki persyaratan penyimpanan yang relatif lebih kecil. Misalnya, gigabita (GB).

Waktu respons : OLTP memiliki waktu respons yang lebih pendek, biasanya dalam milidetik.

Contoh penerapan : OLTP baik digunakan untuk memproses pembayaran, manajemen data pelanggan, dan pemrosesan pesanan.

Data Warehouse

Definisi: Sistem penyimpanan data yang dirancang untuk analisis dan pelaporan, biasanya dari berbagai sistem operasional.

Ciri Khas:

Schema-on-write (struktur data sudah ditentukan saat disimpan)

Data sudah diproses dan dibersihkan (cleaned & transformed)

Ideal untuk OLAP dan BI tools seperti Power BI, Tableau

Contoh Tools: Amazon Redshift, Google BigQuery, Snowflake

Data Lake

Definisi: Tempat penyimpanan besar untuk raw data (data mentah) dari berbagai sumber, bisa berupa structured (tabel), semi-structured (JSON, XML), maupun unstructured (gambar, video).

Ciri Khas:

Schema-on-read (struktur data ditentukan saat data diambil, bukan saat disimpan)

Fleksibel dan murah

Cocok untuk big data, machine learning, dan data science

Contoh Tools: Amazon S3, Azure Data Lake, Hadoop HDFS

Data Mart

Definisi: Subset dari data warehouse yang fokus pada satu departemen atau fungsi bisnis tertentu (misalnya: penjualan, keuangan).

Ciri Khas:

Lebih kecil dan lebih spesifik daripada data warehouse

Akses data lebih cepat untuk user spesifik

Bisa dibuat secara top-down (dari data warehouse) atau bottom-up

Sumber Lain : <https://www.indeed.com/career-advice/career-development/data-lake-vs-data-warehouse-vs-data-mart>